UNIVERSIDADE FEDERAL FLUMINENSE

NOME(S) DO(S) AUTOR(ES)

* ESTA PÁGINA É Capa Dura a ser impressa na gráfica.
* **Com hifenização**.

(remova as caixas amarelas)

TÍTULO DO TRABALHO

Niterói

ANO

NOME(S) DO(S) AUTOR(ES)

TÍTULO DO TRABALHO

Trabalho de Conclusão de Curso submetido ao Curso de Tecnologia em Sistemas de Computação da Universidade Federal Fluminense como requisito parcial para obtenção do título de Tecnólogo em Sistemas de Computação.

Orientador(a):

Nome

NITERÓI

ANO

Folha reservada para a ficha catalográfica

NOME(S) DO(S) AUTOR(ES)

TÍTULO DO TRABALHO

SUBTÍTULO

Trabalho de Conclusão de Curso submetido ao Curso de Tecnologia em Sistemas de Computação da Universidade Federal Fluminense como requisito parcial para obtenção do título de Tecnólogo em Sistemas de Computação.

Niterói, \_\_\_ de \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ de ANO.

Banca Examinadora (provisório):

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Prof. ou Profa. <NOME>, <Título>. – Orientador ou Avaliador

<Sigla da Universidade> - <Nome da Universidade>

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Prof. ou Profa. <NOME>, <Título>. – Orientador ou Avaliador

<Sigla da Universidade> - <Nome da Universidade>

A dedicatória é opcional, no qual o autor presta homenagem ou dedica seu trabalho a alguém.

Dedico este trabalho a(o) minha(meu) esposa(o) e aos meus estimados filhos.

AGRADECIMENTOS

(EXEMPLO) A Deus, que sempre iluminou a minha caminhada.

A meu Orientador Fulano de Tal pelo estímulo e atenção que me concedeu durante o curso.

Aos Colegas de curso pelo incentivo e troca de experiências.

A todos os meus familiares e amigos pelo apoio e colaboração.

Epígrafe é a inscrição colocada no início de um trabalho, de um capítulo ou partes principais. É opcional para a monografia.

Delete as caixas amarelas

“A Escola é uma arena onde grupos sociais lutam por legitimidade e poder”.

Dinair Leal da Hora

RESUMO

O resumo deve apresentar os pontos relevantes de um texto, no mesmo momento ele terá que dar uma visão rápida e clara dos conteúdos, metodologias e das conclusões do trabalho. As frases têm que ter uma sequência lógica e objetiva. A fonte do estilo é Arial e tamanho 12.

Palavras-chaves: palavra1, palavra2 e palavra3.

*O resumo e abstract* devem ser digitados sem parágrafo inicial e o espaçamento entre linhas é simples. Lembre-se que já existem estilos que formatam o texto adequadamente. Remova as caixas amarelas antes de imprimir.

ABSTRACT (opcional)

The summary must present the excellent points of a text, at the same moment it will have that to give a fast and clear vision of the contents, methodologies and of the conclusions of the work. The phrases have that to have a logical and objective sequence. The source of the style is so great Arial and 12.

Key words: word1, word2 and word3.

LISTA DE ILUSTRAÇÕES

[**Figura 1: Exemplo de Figura. 19**](#__RefHeading___Toc378694305)

LISTA DE TABELAS

[Tabela 1: Exemplo de Tabela. 19](#__RefHeading___Toc378694310)

LISTA DE GRÁFICOS

[Gráfico 1: Exemplo de um gráfico 18](#__RefHeading___Toc378694318)

LISTA DE ABREVIATURAS E SIGLAS

LDB – Lei de Diretrizes e Bases

MALO – Museu ao Ar Livre de Orleans

SMA – Secretaria Municipal de Administração

SME – Secretaria Municipal de Educação

SUMÁRIO

[RESUMO 7](#__RefHeading___Toc378694359)

[ABSTRACT (opcional) 8](#__RefHeading___Toc378694360)

[LISTA DE ILUSTRAÇÕES 9](#__RefHeading___Toc378694361)

[LISTA DE TABELAS 10](#__RefHeading___Toc378694363)

[LISTA DE GRÁFICOS 11](#__RefHeading___Toc378694364)

[LISTA DE ABREVIATURAS E SIGLAS 12](#__RefHeading___Toc378694365)

[1 INTRODUÇÃO 15](#__RefHeading___Toc378694366)

[2 DESENVOLVIMENTO DO TRABALHO 16](#__RefHeading___Toc378694367)

[2.1 TEXTO DO TRABALHO 16](#__RefHeading___Toc378694368)

[2.1.1 REFERÊNCIAS BIBLIOGRÁFICAS 17](#__RefHeading___Toc378694369)

[2.1.2 CITAÇÕES 17](#__RefHeading___Toc378694370)

[2.1.3 IDIOMA ESTRANGEIRO 18](#__RefHeading___Toc378694371)

[2.1.4 FIGURAS, GRÁFICOS E TABELAS. 18](#__RefHeading___Toc378694372)

[2.1.5 NOTAS ENTRE O ORIENTADOR E O ALUNO 19](#__RefHeading___Toc378694373)

[CONCLUSÕES E TRABALHOS FUTUROS 20](#__RefHeading___Toc378694374)

[REFERÊNCIAS BIBLIOGRÁFICAS 21](#__RefHeading___Toc378694375)

[ANEXOS 22](#__RefHeading___Toc378694376)

# INTRODUÇÃO

Parte inicial do texto, onde devem constar a delimitação do assunto tratado, o problema ou oportunidade, objetivos da pesquisa, motivação para fazer a pesquisa, métodos usados e resultados encontrados, quando for o caso, e outros elementos necessários para situar o tema do trabalho, bem como a estrutura do documento.

A introdução pode ter a seguinte ordem:

* O que é o trabalho? (1 ou 2 parágrafos)
* Motivação para desenvolvê-lo? (1 parágrafo)
* Métodos usados, quando for o caso (linguagens, protocolos, metodologias, etc.) ( 1 ou 2 parágrafos)
* Organização do trabalho:
  + no Capítulo 2....;
  + o terceiro capítulo é dedicado ...;
  + ...; e
  + finalmente, no Capítulo “tal”, temos as conclusões e indicações para futuros trabalhos.

# Fundamentação teórica

Um artigo da Scientific American divulgou que em 2016 produziu-se tantos dados quanto em toda a história da humanidade até 2015. Isso só é possível devido ao surgimento de novas tecnologias que além de facilitar que os seres humanos exponham todo tipo de informação, também vem conectando diversos tipos de objetos à Internet. Além disso, os instrumentos de medição têm evoluído, ao ponto de as métricas serem cada vez mais precisas e abrangentes.

Neste capítulo, serão apresentadas as categorias de dados mais relevantes, as suas formas de armazenamento, e as técnicas para a coleta e análise de dados.

## Coleta de dados

A coleta de dados é um processo que visa adquirir dados acerca de variáveis da qual se tem interesse. Os dados são utilizados não apenas para tarefas de pesquisa, mas, também, para planejamento, desenvolvimento, experimentações, teste de hipóteses, estudos de caso etc. Ou seja, a coleta de dados é o núcleo de qualquer tipo de pesquisa, seja ele em ciências exatas, físicas, humanas e sociais, marketing, negócios e assim por diante.

Esse processo é feito de maneira estruturada utilizando técnicas específicas de pesquisa. Ainda que, em cada área, os métodos de coleta possuam variações, deve-se ressaltar que a garantia de precisão e veracidade da mesma permanece.

A coleta de dados pode se dar através de três fontes de dados: fonte primária, secundária e terciária.

A fonte primária, também definida como fonte original, são as adquiridas diretamente pelo autor da pesquisa. Como por exemplo, as pesquisas de intenção de voto, realizadas pelo DataFolha, no qual os dados de intenção de voto são coletados diretamente da população, utilizando técnicas de amostragem.

A fonte secundária é, neste ponto, onde os dados já sofreram algum tipo de manipulação, sendo agrupadas e organizadas a fim de facilitar o seu uso. Como exemplos, tem-se os dicionários, enciclopédias, bases de dados etc.

As fontes terciárias são guias para encontrar fontes primárias e secundárias. Por exemplo, guias de leitura, livrarias, catálogos etc.

## Tipos de dados

Uma perspectiva pela qual pode-se entender os dados são as diversas formas pelas quais eles são categorizados. Cada esquema de categorização é um modelo relativo a algum aspecto dos dados. Dentre as diversas formas de categorização de dados, essa seção tratará de dois grandes grupos: quantitativos e qualitativos [33] e primários e secundários [35].

Os dados qualitativos são as características que não possuem valores, ou seja, são dados não numéricos que são definidos por várias categorias que representam uma classificação do indivíduo. Além disso, esses dados, podem ser divididos em nominais ou ordinais.

|  |  |  |
| --- | --- | --- |
| **Tipo** | **Descrição** | **Exemplos** |
| Nominal | Não existe ordenação dentre as categorias | Sexo, cor dos olhos, fumante/não fumante, doente/sadio |
| Ordinal | Existe uma ordenação  entre as categorias | Escolaridade, estágio da doença, mês de observação |

Os dados quantitativos **s**ão dados de valores mensuráveis, ou seja, que podem ser expressados através de valores numéricos, podendo ser de caráter intervalar ou razão, tomados em uma escala métrica definida, podemos classificar as mesmas como discretas quando assumem valores finitos, e contínuas quando assumem valores infinitos.

|  |  |  |  |
| --- | --- | --- | --- |
| **Variáveis** | **Tipo** | **Descrição** | **Exemplos** |
| Quantitativas | Discreta | Conjunto com um valor enumerável. | Número de alunos, Número de filhos, Idade, etc. |
| Continua | Conjunto correspondente a um intervalo real | Peso, Altura, Salário, etc. |

**Pode-se a partir dessa tabela definir a forma de como esses dados podem vir a ser coletado, em seguida processado e apresentados em sua forma final.**

Como mencionado na seção 2.1, os dados primários estão associados as fontes primárias de coleta. Sendo assim, eles são obtidos pelo pesquisador diretamente com o objetivo de abordar o fenômeno estudado. Além do mais, os pesquisadores ao partirem de uma ação planejada, tornam a extração dados mais precisa permitindo responder perguntas direcionadas o que concede maior confiabilidade aos dados coletados.

Em geral, como desvantagem, esse tipo de dado possui alto custo, pois exige a constituição de uma equipe para coleta e sua extração é de difícil acesso. Os dados primários são coletados por meio de diferentes métodos de pesquisa. Como por exemplo: entrevistas em profundidade, questionários, estudos de caso, grupos de discussão etc.

Também, como mencionado na seção 2.1, os dados secundários são associados as fontes secundárias de coleta. São obtidos através de bases de dados como IBGE, FGV, IPEA entre outros. São mais rápidos de obter e possuem um custo menor

Sua desvantagem, em geral, está nos dados não serem atualizados e que nem todas as perguntas a qual são importantes estão respondidas. Por consequência o sucesso da pesquisa pode ser afetado. Logo, é imprescindível a atenção quanto a metodização da coleta, cuidando desde a definição de critérios quanto a seleção de fontes, metodologia empregada para a sua coleta até sua manipulação.

## Organização e armazenamento

Com a coleta de dados houve a necessidade de armazená-los e dispô-los de maneira prática. Os dados podem estar organizados de três maneiras diferentes: estruturados, semiestruturados e não-estruturados.

Os dados estruturados são os mais tradicionais estando presente em sistemas e relatórios. Porém, atualmente, houve um rápido aumento na geração de dados semiestruturados e não-estruturados. Por isso a escolha da forma de armazenamento dependerá da necessidade e do problema [52].

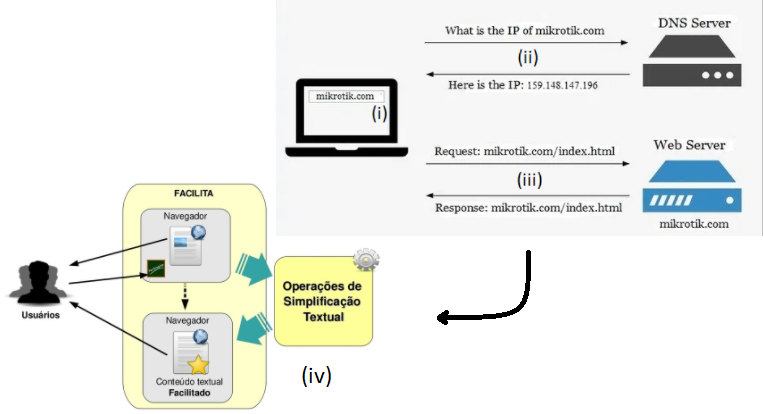
****

Os dados estruturados são dados bem definidos da forma na qual torna-se compreensível e bem definido, seguindo uma pré formatação já estabelecida para a sua utilização, um exemplo de dados estruturados, são banco de dados, onde possuem campos que compõem uma respectiva tabela, organizado em colunas com atributos chaves do tipo: Nome um conjunto de caracteres sem espaços, CPF um conjunto de números não mutáveis, Idade um conjunto de números mutáveis, etc, onde cada linha representando assim um conjunto de informações correspondentes a um atributo principal.

Os dados semiestruturados **s**ão dados que possuem uma estrutura que também é flexível, ou seja, que possuem uma estrutura com um alto grau de alterabilidade, tornando possível qualquer alteração de seu escopo o que permite ter um melhor controle. Um exemplo de normalmente utilizado é o arquivo estruturado XML (*Extensible Markup Lenguage*) que possuem rótulos de abertura e fechamento, possuindo uma forma amigável de visualização para diversos softwares como o Navegador por exemplo.

Os dados não estruturados diferentes dos dados estruturados não possuem um formato pré-definido geralmente são dados que são apresentados de forma bruta, ou seja, são dados que não possuem uma forma ou tamanho máximo, umas forma básica de armazenamento é a partir de um arquivo de testo que pode conter diferentes tipos de dados, outros exemplos de dados não estruturados são: Imagens, Vídeos, Musicas, TXT, etc.

## Navegadores de internet

 Os navegadores com relação a estruturas de páginas, utiliza de páginas semiestruturadas (HTML), de forma a apresentar para o usuário uma interface utilizando as tags para uma formatação mais amigável, conectando-as por através de links que denominados de hipertexto[26], o processo abaixo representa as requisições do acesso ao endereço de um website por um usuário, através de seu navegador, que realiza uma sequência de eventos para a obtenção e o tratamento do HTML que será apresentado ao usuário.

(i) O navegador realiza a extração do nome do hospedeiro do endereço.

(ii)Através do DNS o navegador obtém o endereço de ip do servidor que hospeda a página.

(iii)Por meio do endereço o navegador realiza o download do arquivo HTML, e todos os arquivos correspondentes a referenciação nele presente.

(iv)O navegador realiza operações de simplificação textual para a exibição do conteúdo ao usuário.

## ESTRUTURA DE PÁGINAS

Todos os sites possuem estruturas HTML básica, compostas por uma ou diversas páginas que utilizam de links (hipertexto)[26] para vinculam seus conteúdos, cada página possui seus específicos conteúdos variando entre gráficos, imagens, textos, entre outros formatos.

Cada página possui suas variações, contudo possuem estruturas comuns de nominadas de blocos que são estruturas bem definidas, delimitadas por um par de tags, que definem o seu cabeçalho que contém metainformação sobre seu conteúdo e um corpo que agrupa o restante dos elementos de conteúdos distintos[60].

Na figura a baixo temos a representação de uma página básica, composta da demarcação HTML (Hypertext Markup Language)[27], que foi criado para a internet que foi inicialmente tinha o intuito de facilitar a troca de documentos entre universidades[26], trazendo uma estrutura visual mais clara.



Figura 2: Estrutura básica de um documento HTML

A principal vantagem desse modelo, é a personalização com a possibilidade da introdução de arquivos distintos, tornando um arquivo mais rico em detalhe e consequentemente mais atraente visualmente. Sua popularização ocorreu a partir do ano de 1990[29] com programas capazes de entender e reproduzir o HTML conhecidos como navegadores.

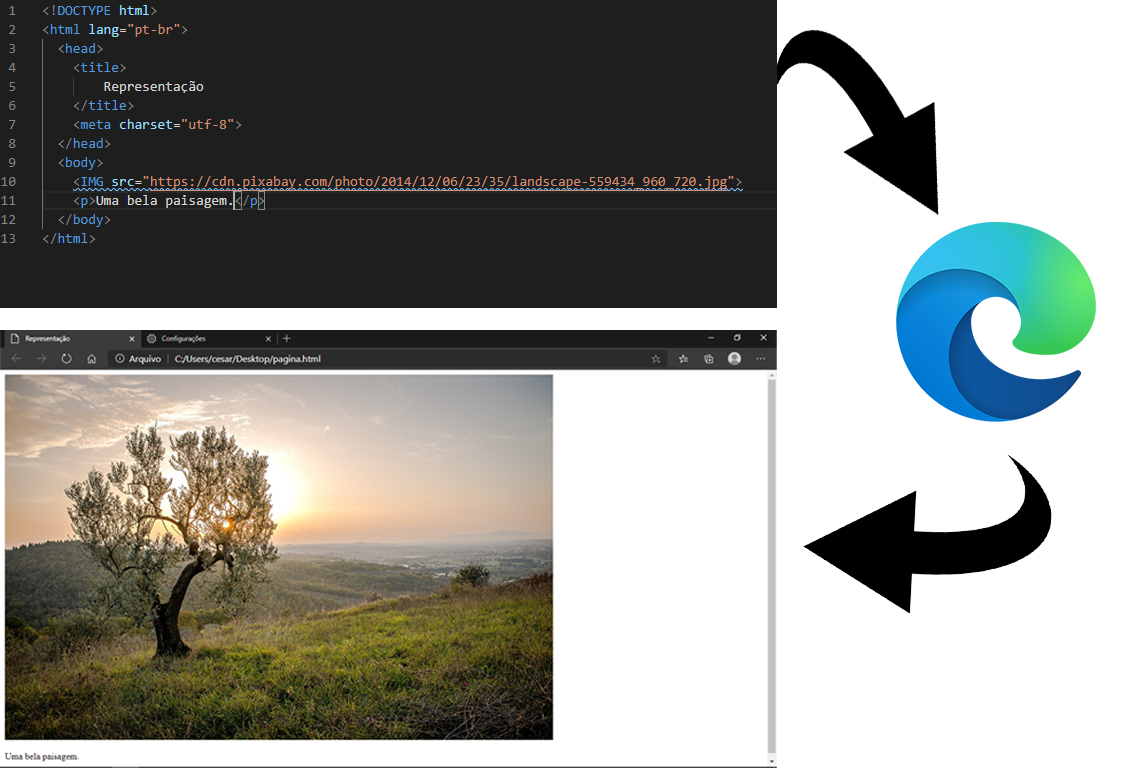


Figura 3: Representação visual do modelo HTML representado em um navegador.

Em 1998 teve implementação do DOM (Document Object Model) que permitiu manipular o HTML como um conjunto de objetos independentes, adicionando um grau de dinamismo as páginas, tornando as em Aplicações de Internet Rica(RIA)[31] que foram suportadas com a evolução dos navegadores. Pode-se destacar nessa nova tecnologia, grupos que tiveram um grande destaque, tais como folhas de estilo(CSS) que flexibiliza a estilização por meio de grupos previamente classificados, JavaScript(JS)[59] que executam códigos de alto nível diretamente no navegador, permitindo alterações em uma página em tempo real.

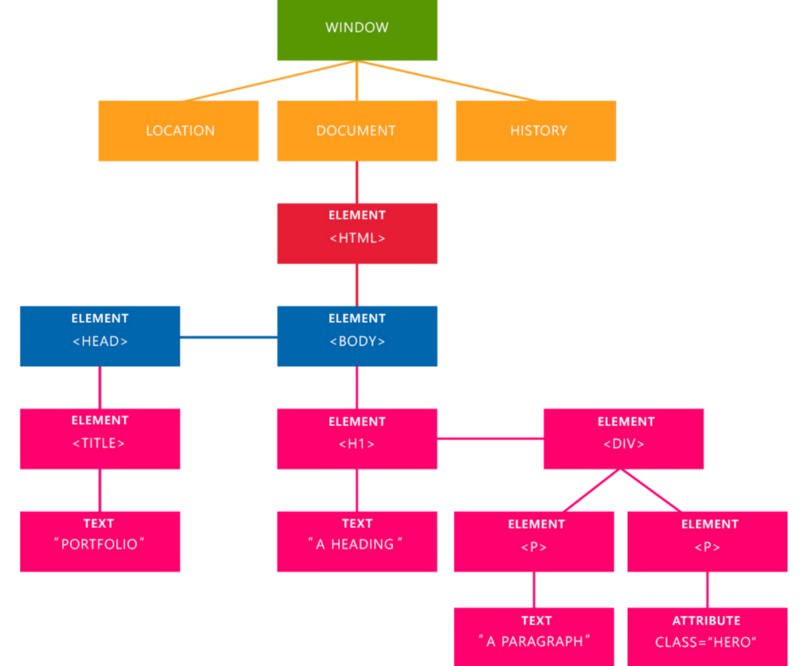


Figura 4: Estrutura do HTML lida através do DOM.

Por mais que o acesso às páginas da web seja mais conhecida por meio de navegadores, há diversos programas de que podem interagir com clientes para os servidores, e utilizando do mesmo princípio apresentado nos navegadores, utilizando de linguagens de alto nível, utilizando de algumas linhas de código para extrair o código HTML de qualquer página, portanto que ele possua o acesso à mesma.

Temos então que uma página da web não se trata apenas de um padrão sólido consolidado, mas um conjunto de elementos definidos que fazem parte de toda internet, que possibilita uma avaliação do conteúdo de uma página da web, uma vez que seu conteúdo hipertexto seja obtido com sucesso e tenha seus objetos devidamente interpretados.

Considerando determinado conjunto de páginas de internet como um aplicativo e, também, que ao serem criadas foram empregadas boas práticas de desenvolvimento de *software, seguindo as definições da linguagem, isso as tornará* estáveis e de simples manutenção. Partindo dessa ideia, para que seja possível a aplicação de filtros nessas páginas, será necessário que:

* Os dados de um mesmo tipo sejam identificados por uma mesma classe ou uma mesma tag HTML,
* O documento HTML possua código essencialmente adequado ao ser construído e que seja feita escolha de *tag* mais apropriada para determinada estruturação de conteúdo.
* Mesmo com o aumento de utilização de *frameworks* para a construção de páginas dinâmicas, a estruturação básica do código HTML ainda se mantém, permitindo que a escolha de *tags* específicas seja feita.

**Portanto, tendo em vista que uma determinada página possua os atributos mencionados acima, a mesma poderá ser escolhida como fonte de extração de uma certa coleção de dados.**

## WEB SCRAPING

No ambiente web, as principais fontes de dados são os documentos HTML. Esses documentos são, a princípio, publicados para serem exibidos em navegadores (web browsers). No entanto, como esses dados são objetos de interesse para estudos, eles devem ser convertidos em estruturas mais apropriadas a tal. O processo de coleta pode ser feito de duas maneiras: manualmente ou automatizado.

* Na modalidade manual, navega-se até a página web desejada onde as informações de interesse serão identificadas visualmente e, por fim, o dado será coletado para processamento posterior.
* Já na modalidade automatizada, utiliza-se um aplicativo de computador para realizar as requisições HTTP em busca de uma determinada página web. Ao receber a página, o aplicativo fará a busca pelas informações demandas e, ao final, retornará as partes do documento no qual o dado está presente.

Como em todo documento de texto, nas páginas web pode-se utilizar técnicas de busca desde as mais simples, como a busca por palavra-chave, até as mais complexas, como a busca por expressões regulares *regex*, do inglês regular expression [40].

No entanto, as técnicas de busca, da qual se tem interesse, são as que utilizam a estrutura do documento para alcançar uma determinada informação. A título de exemplo, tem-se a linguagem Xpath [41] que estabelece um caminho até a informação. Há, também, ferramentas como a biblioteca *Beautiful Soup* [5] que realizam o *parsing* do documento, ou seja, criam uma representação encadeada de objetos.

O *parsing* é a análise sintática do texto de qualquer linguagem de programação, inclusive dos documentos HTML. Sua função é verificar se a estrutura de uma página web está de acordo com as regras que a definem. Portanto, ao encontrar entradas que não estejam dentro das regras gramaticais da linguagem, o analisador sintático as recusará. Em outras palavras, é uma etapa onde ocorrerá a tradução de strings em uma estrutura compreensível pelo computador e que, posteriormente, poder-se-á extrair informações de interesse.

### REFERÊNCIAS BIBLIOGRÁFICAS

Na seção de “**REFERÊNCIAS BIBLIOGRÁFICAS**” temos dois exemplos para o formato das referências. Elas devem ser utilizadas no texto com o formato [1] para a primeira referência da lista e assim por diante, em ordem. Se utilizar um *site* como referência use o modelo em [2]. Quando se referir a uma página específica, como no caso de uma citação use [1, p.32]. Lembre-se de que definições, conceitos, estatísticas, gráficos, ou seja, qualquer afirmação ou dado contido no texto que não for seu deve exibir uma referência. Outra questão é que devemos ter cuidado na seleção das referências, por exemplo, o Wiki não é uma referência confiável, pois existem muitos erros. Uma boa dica é usar a referência automática que já coloca as citações na ordem correta e faz a referência (no final do documento) de forma automática. Com isso, caso alguma referência seja acrescentada depois, você não precisa rever toda a lista citada. Para isso vá no menu referência e clique em inserir citação [1].

### CITAÇÕES

Citação: É a menção do texto de informação extraída de outra fonte para esclarecer, ilustrar ou sustentar o assunto apresentado. Podemos classificá-las em Curta e Longa. “Curta: É transcrita entre aspas, com o mesmo tipo e tamanho da letra utilizados no parágrafo do texto no qual será inserido. O uso das aspas delimita a citação direta”. [1, p.154].

(Exemplo de Citação Longa) É transcrita em parágrafo distinto. Inicia na margem de parágrafo, sem deslocamento na primeira linha e termina na margem direita. Longa: É transcrita em parágrafo distinto. Inicia na margem de parágrafo, sem deslocamento na primeira linha e termina na margem direita. Longa: É transcrita em parágrafo distinto. Inicia na margem de parágrafo, sem deslocamento na primeira linha e termina na margem direita [1, p. 155].

### IDIOMA ESTRANGEIRO

Todos os termos que não pertençam à língua portuguesa devem ser destacados em *itálico*. Os termos não usuais devem ser definidos ou no texto ou em notas de rodapé.

### FIGURAS, GRÁFICOS E TABELAS.

As figuras, gráficos e tabelas devem ser referenciadas no texto, por exemplo: no Gráfico 1 apresentamos um exemplo. Os índices já foram criados neste texto, para colocar novos objetos pressione o botão direito do *mouse* sobre o objeto, selecione “legenda” e digite sua descrição para o objeto. Após este procedimento basta atualizar o índice que ele será incluído.

Gráfico 1: Exemplo de um gráfico

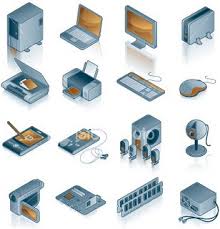


Figura 1: Exemplo de Figura.

Tabela 1: Exemplo de Tabela.

|  |  |
| --- | --- |
| Tabela exemplo | Descrição |
| 1 | Campo 1 |
| 2 | Campo 2 |

### NOTAS ENTRE O ORIENTADOR E O ALUNO

Nossa interação é feita através de e-mails, mas essa forma de comunicação pode ser prejudicial se escrevermos um texto muito longo, portanto devemos ser objetivos. As observações diretamente no texto do TCC têm se demonstrado mais produtivas, a forma e estratégia são combinadas entre o tutor orientador e o orientando.

CONCLUSÕES E TRABALHOS FUTUROS

Neste capítulo devem ser colocadas as conclusões que o aluno obteve durante a elaboração do trabalho, bem como o que pretende após sua conclusão (especialização, mestrado, aplicar os conhecimentos em alguma área...).

REFERÊNCIAS BIBLIOGRÁFICAS

Aqui vale a dica da utilização da forma automática no MSWord ou no LibreOffice:

Neste momento, você começa a elaborar em índice analítico todas as referências que usaram-se para fazer o trabalho monográfico.

1. CARDOSO, Alcionê Damásio. **Vantagens e Desvantagens na Forma de Escolha de Diretor de Escola na Rede Pública Estadual de Santa Catarina**, 2002. Dissertação (Mestrado em Gestão Institucional) – Curso de Pós Graduação em Educação, UnC-UNICAMP, Caçador, SC.
2. SOUZA, Celso de Oliveira. **Histórico da Fundação Educacional Barriga Verde.** <http://www.febave.org.br/historico.htm> Acesso em 25 abr. 2004.

ANEXOS

A função das duas definições, Anexo e Apêndice, é semelhante, mas com uma grande diferença entre elas: a autoria. O ANEXO de um trabalho acadêmico deve ser aquele texto ou documento que não foi elaborado por você, tendo como objetivo servir de legitimação. Já o APÊNDICE se configura como texto ou documento elaborado por você, tendo como objetivo complementar a sua argumentação.

1. ANEXO A – TÍTULO DO ANEXO A

ANEXO B – TÍTULO DO ANEXO B